



Klasifikasi Sentimen Positif dan Negatif pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine (SVM)

Hylmi Andreansyah¹, Sunardi²

^{1,2}Teknik Informatika, Universitas Stikubank, Indonesia

Email: hylmiandreansyah4@gmail.com¹, sunardi@edu.unisbank.ac.id²

Abstract

Gojek is an Indonesian decedent company seeing tremendous growth. Given the growing user base of the Gojek application, it is imperative for the firm to consistently maintain and enhance its services. Sentiment analysis is an efficient method for quickly collecting valuable textual information from a vast quantity of data. The objective of this study is to conduct sentiment analysis on comments from users of the Gojek application on the Google Play Store. The study procedure involves many stages: data gathering, data labeling, data preprocessing, feature extraction, and sentiment classification using the Support Vector Machine (SVM) algorithm. The dataset used consists of user comments from the Gojek application on the Google Play Store, collected via Kaggle, spanning from January to March 2023. This research shows that using the Support Vector Machine method to classify Gojek application user sentiment provides quite high accuracy results, with the highest value in the proportion scenario of 80% training data and 20% test data (accuracy 0.676).

Keyword: Gojek, Analisis sentimen, Support vector machine

Abstrak

Gojek adalah perusahaan dekaden Indonesia yang mengalami pertumbuhan luar biasa. Mengingat basis pengguna aplikasi Gojek yang terus bertambah, penting bagi perusahaan untuk secara konsisten mempertahankan dan meningkatkan layanannya. Analisis sentimen adalah metode efisien untuk mengumpulkan informasi tekstual berharga dengan cepat dari sejumlah besar data. Tujuan dari penelitian ini adalah untuk melaksanakan analisis sentimen terhadap komentar pemakai aplikasi Gojek di Google Play Store. Prosedur penelitian melibatkan banyak tahapan: pengumpulan data, pelabelan data, preprocessing data, ekstraksi fitur, dan klasifikasi sentimen memakai algoritma Support Vector Machine (SVM). Dataset yang digunakan terdiri dari komentar pemakai aplikasi Gojek di Google Play Store, yang dikumpulkan melalui Kaggle, mulai Januari hingga Maret 2023. penelitian ini menunjukkan bahwa penggunaan metode Support Vector Machine untuk mengklasifikasikan sentimen pengguna aplikasi Gojek memberikan hasil akurasi yang cukup tinggi, dengan nilai tertinggi pada skenario proporsi 80% data latih maupun 20% data uji (akurasi 0,676).

Kata kunci: Gojek, Analisis sentimen, Support vector machine

1. PENDAHULUAN

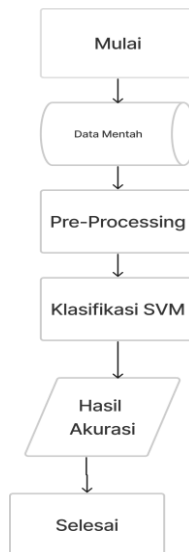
Kemajuan teknologi di Indonesia telah memudahkan berbagai aktivitas manusia yang dapat difasilitasi melalui metode digital, khususnya di industri transportasi. Dinamika ini terlihat dari menjamurnya beragam model transportasi online di wilayah perkotaan besar di Indonesia. Gojek merupakan layanan transportasi online yang sangat populer di Indonesia. PT Aplikasi Karya Anak Bangsa (Gojek) adalah perusahaan decacorn Indonesia berdiri pada 2010 di Jakarta dari Nadiem Makarim. Untuk melaksanakan bisnisnya, Gojek menciptakan aplikasi yang terdaftar di berbagai platform, termasuk Google Play Store, yang terintegrasi dengan *smartphone* [1]. Analisis sentimen digunakan untuk

mengklasifikasikan jawaban pengguna ke dalam kategori positif dan negatif. Saat mengkategorikan ide melalui penambahan teks, penting untuk menggunakan teknik klasifikasi yang tepat maupun akurat. Salah satu caranya adalah metode Support Vector Machine (SVM). SVM memberikan beberapa manfaat, seperti penggunaan fungsi kernel untuk menerapkan hyperplane pada data masukan berdimensi tinggi yang kompleks, sehingga meningkatkan kinerjanya [2].

Penelitian sebelumnya yang dilakukan oleh Diki Muhammad Hendriyanto (2022) berjudul 'Analisis Sentimen Ulasan Aplikasi Mola di Google Play Store' mencapai hasil bahwa dari 520 ulasan, akurasi mencapai 92,31%, presisi 96,3%, recall 89,66%, maupun *f1-score* sebesar 92,86%. Penelitian lain yang dilakukan oleh Fakhri Faizal Irfani (2020) dengan judul 'Analisis Sentimen Review Aplikasi Ruang Guru' menghasilkan akurasi 90,2%, presisi 90,3%, dan *recall* 90,9% dari 2000 data review yang diambil melalui proses *crawling* data.

2. METODOLOGI PENELITIAN

Penelitian ini melibatkan banyak langkah, ialah pengumpulan data dari website Kaggle, pra-pemrosesan data, pelabelan, dan pengembangan sistem klasifikasi memakai algoritma SVM [3].



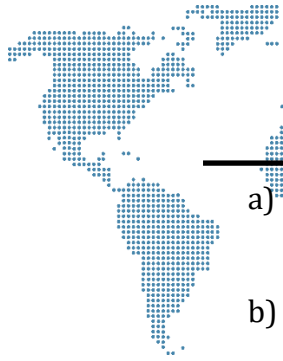
Gambar 1. Flowchart Penelitian

2.1. Pengumpulan Data

Data penelitian ini bersumber dari ulasan aplikasi Gojek pada Google Play store. Pengambilan dataset didapat dari *Kaggle*, dengan jumlah data 20.000 ulasan pemakai dari bulan januari 2023 sampai bulan maret 2023.

2.2. Tahap Pre-processing

Pre-processing adalah langkah pengolahan data teks yang melibatkan penghapusan data yang tidak relevan dalam tahapan analisis sentimen[4]. Dalam penelitian ini, beberapa metode telah diterapkan, antara lain:



- a) Langkah pertama dari *preprocess* teks adalah transformasi, di mana data input mengalami perubahan. Transformasi ini secara default mencakup pengubahan huruf menjadi huruf kecil [5].
- b) *Tokenizing* adalah proses memecah data teks menjadi token individual. Tokenisasi adalah proses membagi suatu teks menjadi kata-kata individual dengan mengenali karakter tertentu yang berfungsi sebagai pemisah kata atau sebaliknya.
- c) *Filtering* adalah langkah menghilangkan kata-kata yang tidak penting setelah proses tokenisasi. Jenis kata tertentu, misal kata keterangan, kata seru, kata ganti, kata depan maupun kata sambung, dapat dikecualikan. Fungsi *regex* digunakan untuk menghilangkan tanda baca, seperti terlihat pada contoh berikut:
“\.,|:|;|!|\?|\(|\)|\|\\+|'|"|"'"|'|\'|...|-|-|—|\\$|&|*|>|”
- d) Normalisasi memiliki fungsi dalam mengubah kata-kata yang tidak baku ataupun tidak sesuai dengan ejaan standar. Normalisasi dilaksanakan memanfaatkan *database* kamus yang memuat kata-kata dalam bentuk baku maupun tidak baku. *Database* ini dibuat didasarkan data komentar sebagai referensi.

2.3. Analisis Sentimen

Analisis sentimen dapat diidentifikasi menjadi penambahan opini ataupun langkah *interpretasi* data teks agar memperoleh informasi. Penggunaan analisis sentimen bisa bermanfaat dalam berbagai konteks, seperti menemukan kelemahan suatu produk, memahami pandangan masyarakat terhadap fitur aplikasi, atau menggali opini terkait suatu topik pembicaraan [6]. Analisis sentimen dipakai dalam mengamati kecondongan opini suatu objek, apakah berupa pendapat positif, negatif ataupun netral [7].

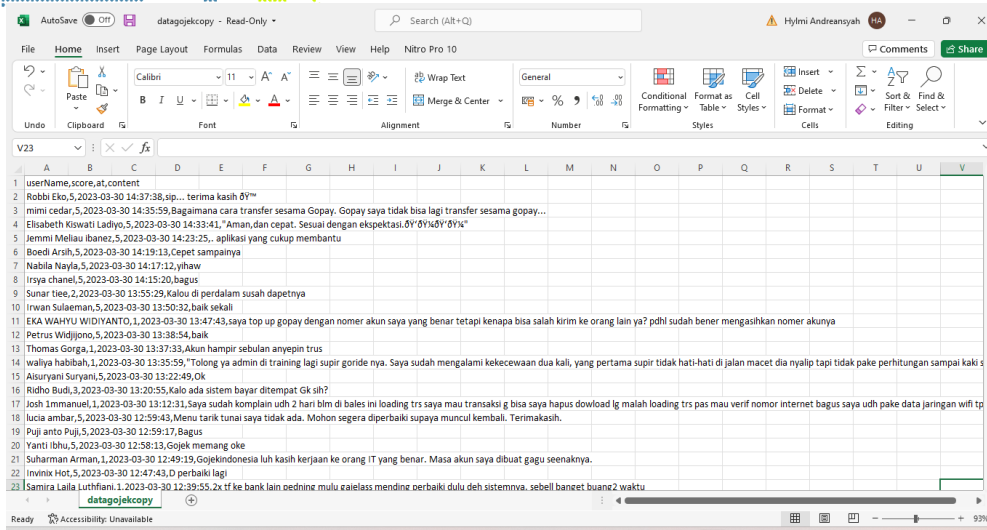
2.4. Support Vector Machine (SVM)

Support Vector Machine (SVM) ialah teknik *regresi* atau klasifikasi data yang memanfaatkan informasi dari data sebelumnya dan mengikuti proses pemodelan yang bersifat *supervisi*. SVM ialah tipe *klasifikator biner* yang bersifat *linier* dan *non-probabilistik*. Dalam penggunaannya, SVM menggunakan *decision boundary* (batas keputusan) untuk melakukan penentuan klasifikasi dari data pelatihan, membentuk model *linier* atau *hyperplane* untuk mengklasifikasikan data [8].

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data pada penelitian ini bersumber dari ulasan aplikasi Gojek pada Google Play store. Pengambilan dataset didapat dari *Kaggle*, dengan total data 20.000 ulasan pengguna dari bulan januari 2023 sampai bulan maret 2023.

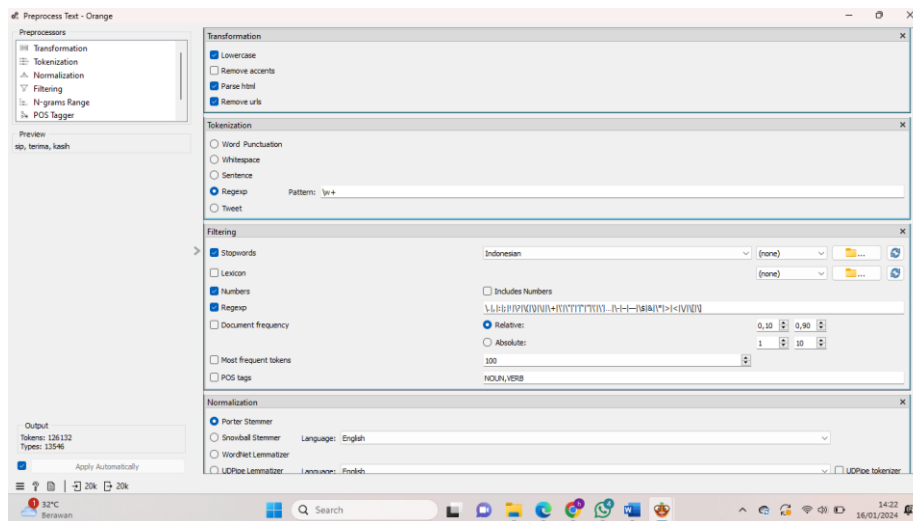


Gambar 2. Hasil Pengambilan Data

3.2. Pre-Processing

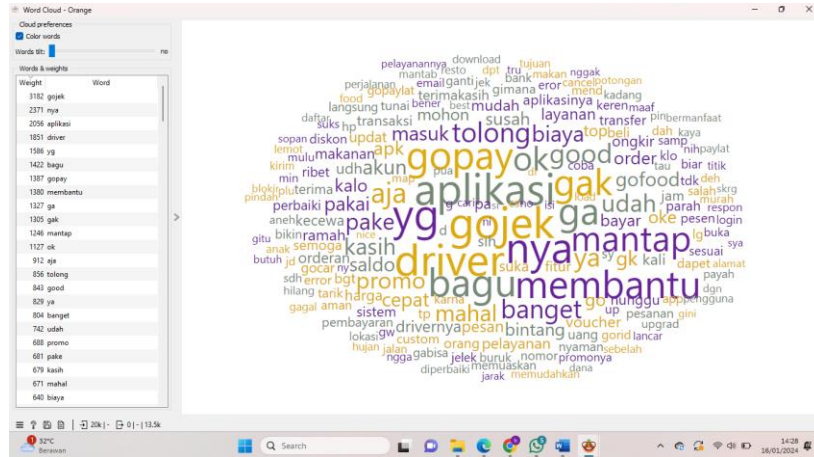
Pre-processing dijalankan menggunakan metode yang telah dipilih sebelumnya, termasuk transformasi, *tokenizing*, *filtering*, dan normalisasi. Transformasi bertujuan mengubah data masukan, dengan transformasi huruf kecil sebagai standar. Proses *tokenizing* kemudian berfungsi untuk membagi karakter-karakter dalam suatu teks menjadi kata-kata, dengan cara mengidentifikasi karakter tertentu yang dapat dianggap menjadi pemisah kata atau sebaliknya.

Selanjutnya, proses *filtering* melibatkan penghapusan kata-kata yang dianggap tidak penting dari langkah *tokenizing* sebelumnya. Kata-kata ini dapat diidentifikasi dan dihilangkan berdasarkan jenisnya, seperti kata keterangan, kata seru, kata ganti, kata depan maupun kata hubung. Selain itu, tanda baca juga dihilangkan dengan menggunakan fungsi *regex*. Tahap terakhir adalah normalisasi, di mana kata-kata yang tidak baku ataupun tidak sesuai ejaan diubah.



Gambar 3. Tahap *Pre-processing*

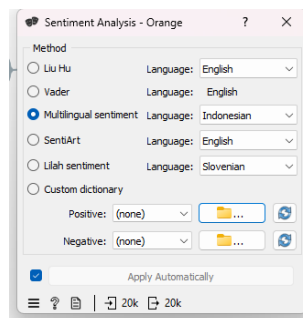
Setelah melalui tahap *pre-processing*, hasilnya berupa data teks yang terpisah, yang bisa divisualisasikan berbentuk *word cloud* menggunakan *platform Orange*. *Widget word cloud* berfungsi untuk menampilkan gambaran visual dengan menonjolkan frekuensi kata yang paling tinggi. Semakin sering suatu kata muncul, ukuran huruf *word cloud* akan semakin besar. Gambar 4 mencerminkan hasil dari representasi visual dalam bentuk *word cloud*.



Gambar 4. Hasil Word Cloud

3.3. Analisis Sentimen

Proses analisis melibatkan penggunaan algoritma *VADER* untuk mengklasifikasikan *polaritas* (sentimen) dalam kategori positif dan negatif, dengan skor total disebut *compound*. Pada tahap ini, data teks dengan tahap *pre-processing* akan dianalisis menggunakan *VADER*, yang menggunakan kamus *lexicon VADER* terkoneksi dengan server *NLTK (Natural Language Toolkit)*. Setelah memilih *VADER* sebagai teknik analisis, algoritma ini akan mengelompokkan dan memberikan skor pada teks didasarkan nilai tiap kata pada *lexicon VADER*. Hasil akhir dari penilaian ini ialah skor total, yaitu *compound*. Skor *compound* ini akan direkapitulasi maupun dibandingkan. Dalam proses analisis, terdapat variabel atribut yang ditekankan seperti *pos*, *neg*, *neu*, dan *compound*, yang menjadi fokus utama *output* analisis. Pengguna dapat memilih atribut-atribut ini menggunakan *widget select columns* untuk menentukan data mana yang dibutuhkan maupun mana yang tidak.



Gambar 5. Pengaturan Analisis Sentimen

3.4. Klasifikasi SVM

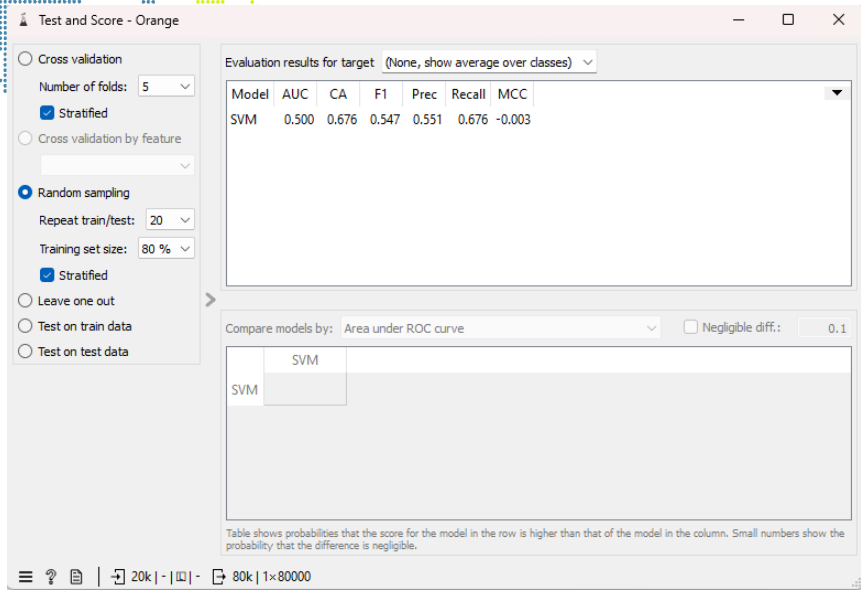
Analisis sentimen dilakukan pada data berlabel yang telah disimpan dalam format xls dan diimpor menggunakan widget corpus di Orange Data Mining. Selanjutnya gunakan fitur pilih kolom untuk hanya melakukan pemilihan kolom Sentimen maupun Konten saja yang akan menjadi landasan penelitian selanjutnya. Tabel data dipakai dalam menyediakan dataset berdasarkan kolom yang dipilih, seperti Gambar 6.

id	title	userName	content	score	at	sentiment	analyst sentiment
1		Robbi Eko	sip... terima kas...	5	2023-03-30 14:3...	66.6667	positif
2		mimi cedar	Bagaimana car...	5	2023-03-30 14:3...	0	negatif
3		Elisabeth Kiswa...	Aman,dan cepa...	5	2023-03-30 14:3...	75	positif
4		Jemmi Meliau i...	. aplikasi yang c...	5	2023-03-30 14:2...	50	positif
5		Boedi Arshih	Cepet sampainya	5	2023-03-30 14:1...	0	negatif
6		Nabilia Nayla	yihaw	5	2023-03-30 14:1...	0	negatif
7		Irya Chanel	bagus	5	2023-03-30 14:1...	0	negatif
8		Sunat tee	Kalou di perdal...	2	2023-03-30 13:5...	-25	negatif
9		Inwan Suleman	baik sekali	5	2023-03-30 13:5...	0	negatif
10		EKA WAHYU W...	saya top up go...	1	2023-03-30 13:4...	0	negatif
11		Petrus Widjijono	baik	5	2023-03-30 13:3...	0	negatif
12		Thomas Gorga	Akun hampir se...	1	2023-03-30 13:3...	0	negatif
13		walya habibah	Tolong ya admi...	1	2023-03-30 13:3...	-5.88235	negatif
14		Aisuryani Suryani	Ok	5	2023-03-30 13:2...	0	negatif
15		Ridho Budi	Kalo ada sistem...	3	2023-03-30 13:2...	0	negatif
16		Josh Immanuel	Saya sudah ko...	1	2023-03-30 13:1...	0	negatif
17		lucia ambar	Menu tarik tun...	5	2023-03-30 12:5...	-14.2857	negatif
18		Puji anto Puji	Bagus	5	2023-03-30 12:5...	0	negatif
19		Nanti Ibtu	Gojek memang...	5	2023-03-30 12:5...	0	negatif
20		Suharman Arman	GojekIndonesia...	1	2023-03-30 12:4...	12.5	positif
21		Invinix Hot	D perbaikan lagi	5	2023-03-30 12:4...	0	negatif
22		Samira Laila Lut...	2x tf ke bank lat...	1	2023-03-30 12:3...	0	negatif
23		Hadisaputra ch...	sangat memba...	5	2023-03-30 12:1...	50	positif
24		Fernanda Cahya	Untuk user gop...	1	2023-03-30 12:0...	0	negatif
25		Amoy Diva Naura	Bagus	5	2023-03-30 12:0...	0	negatif
26		JV bee	Kenapa ya kok ...	2	2023-03-30 12:0...	0	negatif
27		Raden Mas Taka...	Payah	1	2023-03-30 11:5...	-100	negatif
28		RUSIDI (Ross)	Harusnya di da...	1	2023-03-30 11:4...	6.33333	positif
29		ayu destalia	Banyak driver n...	1	2023-03-30 11:3...	20	positif
30		Lia salon	Driver di putus ...	1	2023-03-30 11:2...	-10	negatif
31		Gudeg Celer R...	ini gimana sih ...	1	2023-03-30 11:1...	0	negatif
32		ipey yulanda	divernya oke g...	5	2023-03-30 11:1...	25	positif
33		Endro Wati	Gojek sangat m...	5	2023-03-30 10:5...	33.3333	positif

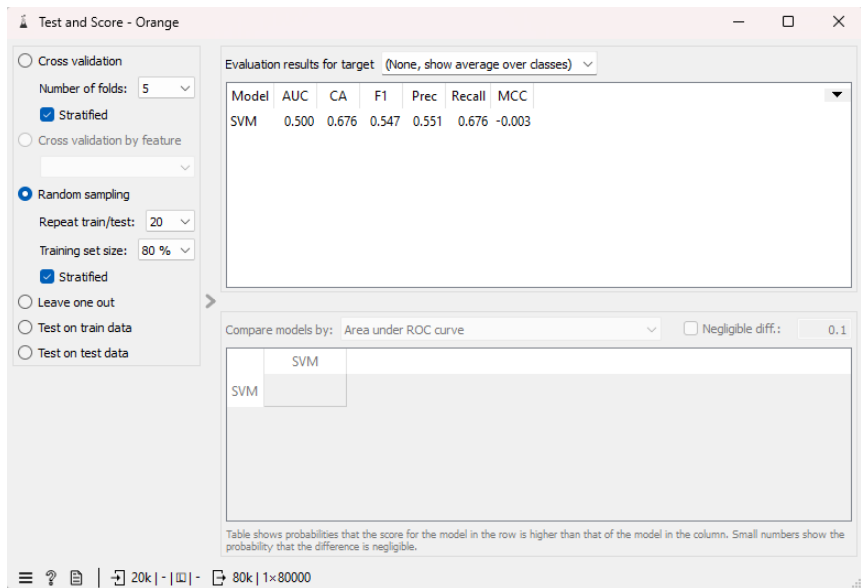
Gambar 6. Data tabel

Kemudian, data sentimen tersebut menjalani proses *pre-processing* seperti Sebelumnya, hal ini tidak diperlukan, namun saat ini penting untuk memiliki komponen filter token, yang sering digunakan untuk membatasi kata-kata yang dianggap penting untuk diproses. Selanjutnya, teknik kantong kata digunakan dalam pengubah data tekstual ke vektor yang diinterpretasikan komputer. Metode ini memfasilitasi penghitungan frekuensi kata di banyak halaman. Frekuensi kemunculan setiap istilah dihitung, dan probabilitas kemunculan klasifikasi positif dan negatif disediakan. Untuk menghasilkan prediksi, seseorang dapat melakukannya dengan membandingkan setiap kata dengan data pengujian dan menghitung probabilitas setiap kata dalam korpus. Peluang masing-masing kelompok akan dibandingkan, dan keputusan diambil berdasarkan kelas yang memiliki probabilitas tertinggi, sehingga dapat diklasifikasikan sebagai positif atau negatif.

Proses selanjutnya adalah tahapan *training*, *testing* menggunakan model SVM. Tahap *training* dan *testing* akan dilakukan 2 kali. Dengan dengan perbandingan data latih 90%, data uji 10% dan perbandingan data latih 80%, data uji 20%. Selanjutnya *widget test and score* digunakan untuk melakukan proses klasifikasi. Seperti yang bisa dilihat pada gambar 7 dan gambar 8 adalah hasil dari proses klasifikasi yang telah dilakukan.



Gambar 7. Hasil Pengujian 1



Gambar 8. Hasil Pengujian 2

Dari kedua hasil pengujian tersebut dapat dilihat pada tabel di bawah ini:

Tabel 1. Perbandingan hasil pengujian

(Data latih, Data uji)	<i>Acuraccy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
80%, 20%	0,676	0,551	0,676	0,547
90%, 10%	0,675	0,532	0,675	0,546

4. SIMPULAN

Kesimpulan dari penelitian ini memperlihatkan pemakai metode *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen pengguna aplikasi Gojek memberikan hasil akurasi yang cukup tinggi, dengan nilai tertinggi pada skenario proporsi 80% data latih maupun 20% data uji (akurasi 0,676). Meskipun demikian, nilai presisi, *recall* maupun *F1-score* masih dapat ditingkatkan. Dalam melihat prospek pengembangan, penelitian selanjutnya dapat fokus pada optimasi model SVM, peningkatan ukuran dataset, eksplorasi ekstraksi fitur yang lebih kompleks, analisis sentimen yang lebih mendalam, dan aplikasi lebih lanjut pada skala yang lebih besar. Peningkatan dalam aspek-aspek ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan sistem analisis sentimen yang lebih akurat dan relevan untuk meningkatkan kualitas layanan aplikasi seperti Gojek di masa depan.

DAFTAR PUSTAKA

- [1] M. N. Muttaqin And I. Kharisudin, "Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine Dan K Nearest Neighbor," *Unnes J. Math.*, Vol. 10, No. 2, Pp. 22–27, 2021, [Online]. Available: [Http://Journal.Unnes.Ac.Id/Sju/Index.Php/Ujm](http://journal.unnes.ac.id/sju/index.php/ujm)
- [2] L. H. Vina Fitriyana, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," *Jatiji (Jurnal Tek. Inform. Dan Sist. Informasi)*, Vol. 9, No. 4, Pp. 3436–3442, 2022, Doi: 10.35957/Jatiji.V9i4.3586.
- [3] A. J. Putri, A. S. Syafira, M. E. Purbaya, And D. Purnomo, "Analisis Sentimen E-Commerce Lazada Pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine," *J. Trinistik J. Tek. Ind. Bisnis Digit. Dan Tek. Logistik*, Vol. 1, No. 1, Pp. 16–21, 2022, Doi: 10.20895/Trinistik.V1i1.447.
- [4] D. Diandra Audiansyah, D. Eka Ratnawati, And B. Trias Hanggara, "Analisis Sentimen Aplikasi Myxl Menggunakan Metode Support Vectormachine Berdasarkan Ulasan Pengguna Di Google Play Store," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 6, No. 8, Pp. 3987–3994, 2022, [Online]. Available: [Http://J-Ptiik.Ub.Ac.Id](http://j-ptiik.ub.ac.id)
- [5] Ismail, H. N. Rahmah, And R. Sulistiyowati, "Penggunaan Software Orange Data Mining Pada Implementasi Text Mining Dalam Analisis Sentimen Netizen Di Twitter Terhadap Kelangkaan Minyak Goreng," *Sigma-Mu*, Vol. 14, No. 2, Pp. 1–11, 2022, Doi: 10.35313/Sigmamu.V14i2.4667.
- [6] Raksaka Indra Alhaqq, I Made Kurniawan Putra, And Yova Ruldeviyani, "Analisis Sentimen Terhadap Penggunaan Aplikasi Mysapk Bkn Di Google Play Store," *J. Nas. Tek. Elektro Dan Teknol. Inf.*, Vol. 11, No. 2, Pp. 105–113, 2022, Doi: 10.22146/Jnteti.V11i2.3528.
- [7] B. Pamungkas, M. E. Purbaya, And D. J. A.K, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (Svm) Pada Kasus Benih Lobster 2020," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, Vol. 3, No. 2, Pp. 10–20, 2021.
- [8] L. Mutawalli, M. T. A. Zaen, And W. Bagye, "Klasifikasi Teks Sosial Media Twitter Menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto)," *J. Inform. Dan Rekayasa Elektron.*, Vol. 2, No. 2, P. 43, 2019, Doi: 10.36595/Jire.V2i2.117.